

## Identifying factors in estimation of Body Mass Index: cohesion between environmental factors and physical activity

Md. Nofayer Haque, Muhammad Mahady Hasan, Azizur Rahman

Department of Statistics, Jahangirnagar University, Savar, Bangladesh

### Article Info

#### Article history:

Received Jun 9, 2019

Revised Jul 20, 2019

Accepted Aug 21, 2019

#### Keywords:

Artificial Neural Network

Body Mass Index (BMI)

Environmental factor

Multilayer Perceptron

Physical factor

### ABSTRACT

Physical health is the key issue of active living in population. Indicator like Body Mass Index (BMI) is affected by physical activity of individuals in their daily life. Overall health condition in a population can be measured by BMI. The aim of this study was to identify the effects of different physical and environmental factors in estimation of BMI level and fit a suitable ANN based model. The mean BMI of 75 adults of our data was  $23.8671 \pm 2.89217$  kg/m<sup>2</sup>. At first, we obtained the correlation structure of the variables and the perform regression analysis for the extraction of the variables that may significantly influenced the BMI. The statistical analysis showed that despite the apparent association of BMI with physical activity level, it was also influenced by several factors such as age, distance to bus stop, dieting, marital status, educational level and recreational facilities. It shows that daily exercise, recreational facility, education level has significant impact in estimation of Body Mass Index (BMI) and we use these factors as an input in ANN model for the estimation of BMI. Multilayer Perceptron (RP) performed better than Radial Basis Function (RBF) of ANN model in terms of model selection criteria.

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

### Corresponding Author:

Azizur Rahman,  
Department of Statistics,  
Jahangirnagar University, Savar, Dhaka-1342, Bangladesh.  
Email: azizur@juniv.edu

## 1. INTRODUCTION

As an health issue, the Body Mass Index (BMI) has been adjuvant in population-based studies on the account of its wide acknowledgement in ascertaining specific categories. Body Mass Index (BMI) is the metric in use for defining anthropometric characteristics in adults based on their weigh and height. It represents an index of an individual's fatness. Also, it is envisaged as a risk factor for the upliftment of or the prevalence of several health issues. Besides, it is widely used to determine public health policies.

A high BMI can be a symbol of hypersecretion of fat on the body, while a low BMI can be a symbol of mediocre amount of fat on the body. If a person's BMI is higher, it means he has a greater chance of developing certain serious conditions, such as heart disease, high blood pressure, and diabetes. A very low BMI can also cause health problems like bone loss, decreased immune function, and anemia [1-4]. Adults age 20 and older can expound their BMI based on the following standard categories. These are the same both for men and women of all ages and body types [5-8],

BMI	Weight Status
Below 18.5	Underweight
18.5–24.9	Normal
25.0–29.9	Overweight
30.0 and above	Obese

For urban planning one of the critical issues is how to provide the citizens a proper physical activity which is welcomed. In this manner, it is recommended that neighborhood design must be based on walking distance criteria. Physical activity of individuals attributed by different factors such as age, gender, attitude towards exercise, educational and marital status [9]. On the other hand, the Body Mass Index (BMI) level is related to the level of physical activity. Standard way of BMI calculation depends on person's weight in kilograms and person's height squares measured in meters i.e.,  $BMI = \text{weight}/\text{height}^2$  [10]. New insights in health-oriented urban planning can be generated by generalization physical impact on the BMI. Recently, one study revealed that metropolitan area like in Atlanta, daily time spent for physical activity decline on nearest living area and opportunities for recreation facilities [9]. Moreover, a study conducted in New South Wales, Australia based on 1200 adults aged between 40 and 60 years showed that individual who had the opportunity to live in an environment where physical activity instrument is available, they had significant intensity in walking pattern and other physical activity. In [11-12] it was observed that perception of towards physical environment had a significant change in physical activities, indeed individuals who had negative perception of environment were less involved in physical activities. A research in Santiago, California signified that individuals having higher levels of physical activity had lower BMI [13]. The physical activity level was associated with the number of nearby recreational facilities and parks.[14]. So, it can be said that the normal Body Mass Index (BMI) is a result of higher levels of physical activity.

## 2. METHODS AND MATERIALS

As a data collection procedure this study used face to face questionnaire method and a cross-sectional study design had been considered. Our study population was people living in Jahangirnagar University campus area. Study location was selected purposely due to cost and time limitations. During the survey, respondent's age was between 17-67 years. We selected 75 people randomly and use a direct interview method for collecting information [15]. Variables in the questionnaire covered three areas: socio-demographic and economic, environmental and physical activity. Among these three areas two themes were the major contributing factors for BMI. Socio-demographic and economic variables included age, height, income, marital status. Physical activity variables were indoor activities, outdoor activities, daily exercise. The data processing procedures includes editing the questionnaire, coding, classification, tabulation. All analysis of the collected data was performed based on SPSS software.

## 3. EMPIRICAL RESULTS

To begin our analysis, we have to describe our population and this section is devoted to find descriptive statistics of our survey data. The distribution of male and female were not equally distributed and about 28 percent were women and 72 percent were male. Also, in our study we had 21.3 percent individuals who had smoking status. In terms of dietary information of individuals, we had 46.7 percent on diet. We are on university area so about 98 percent are educated, 92 percent just finished high school, 81.3 percent had higher school certificate, 49 percent had graduate and 22.7 percent had master's degree and over. Family income aspect, about 24 percent had monthly income less than 28,000 Taka; about 50 percent people income was 28,000 to 48,000. Table 1 depicted descriptive statistics of some selected variables.

Table 1. Descriptive statistics

	Minimum	Maximum	Mean	Std. Deviation
Age of the respondent	17	67	36.00	13.648
Daily Exercise(min)	0	120	20.11	22.102
Distance from the bus stop	.25	8.00	1.4847	1.38513
Body Mass Index (BMI)	15.73	32.02	23.8671	2.89217
Height of the respondent	58	72	64.40	2.809
Weight of the respondent	39	85	63.99	9.430
Income of the family of respondent	8000	100000	38120.00	16684.643

From the Table 1 we see that maximum age of the respondent is 67 and minimum age is 17 average age is 36 and standard deviation is 13.64. Average amount of daily exercise is 20.11, and standard deviation is 22.10. Average distance from the nearest bus stop is 1.5 km and here the standard deviation is 1.38 Minimum Body Mass Index is 15.73, Maximum is 32.02, Average BMI is 23.86, and here the standard deviation is 2.89 Maximum family income of the respondent is 100,000 (TK) and minimum family income is 8,000 (TK) average income is about 38,000 (TK).

We found that only 4% respondent are underweight, 54% are Normal, 39% are Overweight and rest 3% are Obese. The connection between Body Mass Index and Marital Status of the respondent that 55 percent respondents are normal among them 53 percent are male, and 38 percent respondents are overweight among them 63 percent are male. The connection between Body Mass Index and Marital Status of the respondent represents that 41 respondents are normal among them 25 are married, 29 respondents are overweight among them 23 are married. The result of Pearson Chi-Square test can be seen in Table 2.

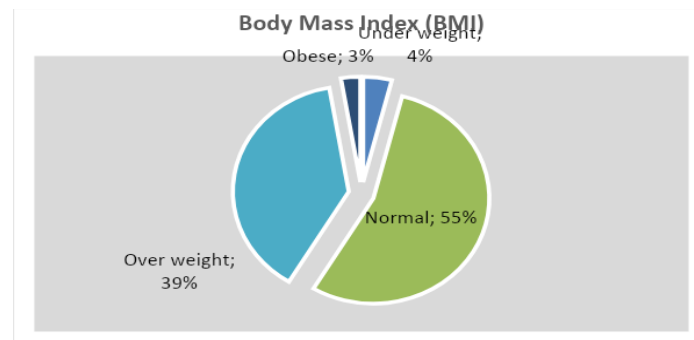


Figure 1. Body Mass Index (BMI)

Table 2. The result of Pearson Chi-square test

	Pearson Chi-Square	df	Asymp. Sig. (2-sided)
Gender	9.065	3	.028
Marital Status	9.684 <sup>a</sup>	3	.021

Gender and Marital status of the respondent has a significant association with body mass index. Moreover, to examine whether there is any difference in male and female group BMI controlling the effects of smoking and dietary status we conducted an independent sample t-test. The result revealed that BMI was significantly different between men and women after controlling smoking and dietary status. The distribution of Age, height & weight according to gender is depicted in Table 3.

Table 3. Age, height & weight according to gender

		Range	Minimum	Maximum	Mean	Std. Deviation
Female	Age	30	21	51	32.71	9.46
	Height	7	58	65	61.57	1.85
	Weight	43	39	82	56.90	10.57
Male	Age	50	17	67	37.37	14.84
	Height	12	60	72	65.5	2.3
	Weight	34	51	85	66.74	7.3

We found that average age of female is 32.71 where male is 37.37. And the standard deviation is 9.46 and 14.84. The average height of female is 61.57(inch) where male is 65.5(inch) and the standard deviation is 1.85 and 2.3. The average weight of female is 56.9(kg) where male is 66.74(kg) the standard deviation is 10.57 and 7.3.

Correlation structure in Table 4 indicates that there is a positive correlation among individual's age and weight. On the contrary, all types of physical activity showed negative correlation with BMI [16]. This means that individuals who spent more time on physical functions, they will have less BMI level. A linear regression model was applied to examine the effects of gender, recreational facilities, dieting, education level, smoking habit, marital status of the respondent, distance from bus stop, daily exercise. Hence, Table 5 presents the model summary.

Table 4. Bivariate correlation with Body Mass Index (BMI)

	Age	Distance from the bus stop	Daily exercise	Height	Weight	Body Mass Index
Age	1					
Distance from the bus stop	-0.11119	1				
daily exercise	0.199353	0.061411	1			
height	0.234384	0.087725	0.195835	1		

weight	0.258917	0.029624	0.069185	0.58936	1	
Body Mass Index	0.162193	-0.02297	-0.05543	0.01133	0.81149	1

Table 5. Model summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.606	.367	.291	2.43582

In Table 5 the overall model summary is given. About 29% of total variability is explained by the chosen model which is not a good fitted model. This may happen due to multicollinearity problem. However, our main focus is on estimation of BMI, therefore we ignore the violation of multicollinearity issues in this paper. The result of coefficients is deployed in Table 6.

Table 6. Coefficients

Model	B	Std. Error	t	Sig.
(Constant)	26.497	1.303	20.340	.000
Gender	1.442	.693	2.081	.041
Marital status	-2.754	.611	-4.506	.000
Distance from the bus stop	.299	.226	1.324	.190
Smoking habit	.537	.734	.731	.467
Dieting	-.869	.624	-1.393	.168
Education level	-.669	.246	-2.718	.008
Daily exercise	-.040	.015	-2.649	.010
Recreational facilities	-.716	.253	-2.832	.006

Table 6 provides that the significance level for Gender, Marital Status, Education level, Daily exercises, Recreational facilities are less than .05. So, we can conclude that there is very strong vindication to reject the null hypothesis for the above variables. Therefore, each variable has predictive ability for the dependent variable Body Mass Index (BMI).

Fitted Regression Model is:  $\text{Body Mass Index (BMI)} = 26.497 + 1.442 * \text{Gender} + (-2.754) * \text{Marital Status} + (-.669) * \text{Education level} + (-.040) * \text{Daily exercise} + (-.716) * \text{Recreational facilities}$ . Above model tells that the intercept term is 26.497. If there is no factor exist, then initial BMI will be 26.497. If one unit increasing in daily exercise on an average Body Mass Index (BMI) decreases 0.040 unit if other remains constant.

### 3.1. BMI estimator based on Artificial neural network

ANN is a computational tool which is constructed with some interconnected principal units called Neurons as shown in Figure 2. The most important applications of the Feed Forward Artificial Neural Networks (FFANN) is Recognition Patterns. There are many types of FFANN. Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) are one of the important types, Picking the most compatible structure for ANN and thanks to nonlinear character it can adapt to imminent even such complex patterns—between some impact factors and targets—that customary models never deliver [17]. In Figure 2 we showed ANN scheme:

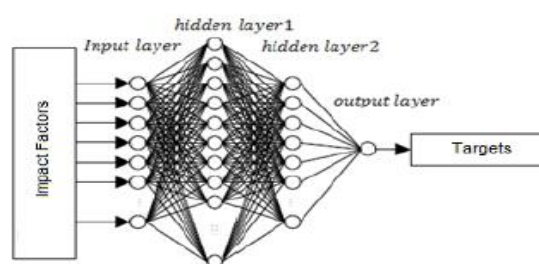


Figure 2. The ANN schemes

### 3.2. Using relationship between BMI and impact factors on training process

Artificial Neural Network (ANN) is used to generate a model envisaging relationships between some input and output data based on training data. Appropriate use of ANN model depends on optimum separation of data into train and test set. Moreover, data set must be representative of all scenarios possible with low superfluity as well as with suitable dimensions. First approach is to train the model with training data set.

Yet, ANN has high potential in approximation in complex patterns, it may flunk in forecasting the patterns with low correlation between input and output. Therefore, in this study, the task was picking the data set which reflected the goal function. So, the cabalistic link between BMI and various environmental factors, physical activity and socio-demographic-based factors suggests a cabbalistic pattern to learn the ANN model on the base of statistical correlation between different influential factors and BMI. The input list for training ANN model is shown in Table 7.

Table 7. Input variable List to train ANN model

Input No.	Significant impact factors
1	Gender
2	Age
3	Marital Status
4	Distance from the bus stop
5	Facility of indoor activities
6	Facility of outdoor activities
7	Recreational Facilities
8	smoking habit
9	Diet
10	Daily exercise
11	Education level

### 3.3. Evaluation of Neural Network Model

The major concern in ANN model implementation is to select the best structure under which model reveals appropriate estimation with lowest number of layers and neurons.[17]. We are going to use ANN model for estimation purpose. Therefore, we developed ANN architecture for two different training algorithms and one Radial Basis Function (RBF) based network. Levenburg-Marguardt (LM) technique and Resilient back Propagation (RP) techniques performed in the training process of ANN. To reach the best estimation results, network structures and training technique were picked by using trial and error approach. Evaluating the trained ANN within new scenarios never used before to test the precision of model on pivotal estimation task which is the most significant step of implementation. The ability of the trained neural network to measure generalization under new scenarios that neural fault locator never experienced before are evaluated at this stage. New scenarios are coherently sampled randomly within the total sample. For a better judgment, we included five criteria such as Mean Value, Standard Deviation (SD), Normalized Root Mean Square of errors (NRMSE), Correlation co-efficient, Maximum Absolute of Error (MAE). The normalized Root Mean Square of Errors is defined as follows:

$$\text{NRMSE} = \sqrt{\frac{\sum (Y(n) - T(n))^2}{\sum T(n)^2}}$$

Here R values in Figure 3 (a) and (b) discloses how well fitted data and target value matched each other. The resulting value nearer to 1 indicates there is a high correlation between T and Y exist. Linear fit also illustrates the exactness of the estimation therefore zero crossing line with the slope of 1. Figure 3 (a) and (b) visually discloses the execution of MLP (RP) and RBF respectively under the new scenarios. From linear fit and R value, we can observe that RBF whose MAE=9.7389 and NRMSE=0.198 yields a less appeasement results in comparison with the MLP-based networks.

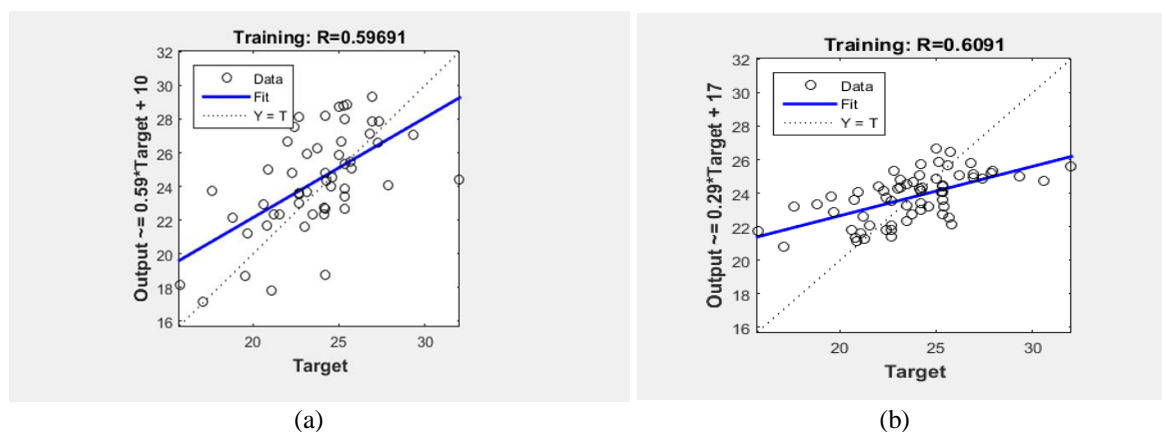


Figure 3. Training process MLP (Multi Layer Perception)

(a) RP (Resilient back propagation), (b) RBF

In Table 8, we depicted the performance comparison output of our two ANN models. The model selection criteria are NRMSE and MAE. For MLP (RP) model we have comparatively lower value of NRMSE and MAE than RBF model. Moreover, the last column of table 8 showed that we have good fit of MLP(RP) model compared to RBF. These indexes are comprehensively influenced by genetic traits, besides, some external factors such as socio-economic, physical activity and environmental-based factors have significant influence on BMI.

Table 8. The performance comparison output

Net. Type	NRMSE	MAE	Mean	Std.	R
MLP(RP)	0.0971	1.84784	23.8424	6.54139	.91789
RBF	0.198	9.7389	23.7372	5.0347	.6091

#### 4. CONCLUSION

Physical activity and environmental factors have significant effect in estimation of BMI, yet it was dependent on some socio-demographic factors as well. From the results section, we observe that BMI was also influenced by several factors such as age, marital status, distance to bus stop, dieting, daily, educational level and recreational facilities. It shows that daily exercise, recreational facility, education level has significant impact in estimation of Body Mass Index (BMI). Moreover, when we attempt to fit ANN based model using MLP and RBF, our model selection criteria support MLP (RP) model as better fit compared to RBF using the input sets from Table 7 presented in section 3.2 of empirical result section. The results of MLP-based ANN provide the significant influence of input factors which can be reasonably be generalized to new scenarios with similar problems.

#### REFERENCES

- [1] Grabisch, M.; Marichal, J.L.; Mesiar, R.; Pap, E. (2009). *Aggregation Functions*. Oxford University Press.
- [2] Jeremy Singer-Vine (July 20, 2009). "Beyond BMI: Why doctors won't stop using an outdated measure for obesity". *Slate.com*. Archived from the original on 7 September 2011. Retrieved 15 December 2013.
- [3] Keys, Ancel; Fidanza, Flaminio; Karvonen, Martti J.; Kimura, Noboru; Taylor, Henry L. (1972). "Indices of relative weight and obesity". *Journal of Chronic Diseases*. 25 (6–7): 329–43. PMID 4650929. doi:10.1016/0021-9681(72)90027-6.
- [4] Malcolm Kendrick (April 12, 2015). "Why being 'overweight' means you live longer: The way scientists twist the facts". <https://www.independent.co.uk>. Archived from the original on 12 April 2015. Retrieved 12 April 2015.
- [5] Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies". *The Lancet*. 363 (9403): 157–63. 2004. ISSN 0140-6736. PMID 14726171. doi:10.1016/S0140-6736(03)15268-3
- [6] Blackburn, Henry; Jacobs, David (2014-06-01). "Commentary: Origins and evolution of body mass index (BMI): continuing saga". *International Journal of Epidemiology*. 43 (3): 665–669. ISSN 0300-5771. doi:10.1093/ije/dyu061.
- [7] "Executive Summary". *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report*. National Heart, Lung, and Blood Institute. September 1998. xi–xxx. Archived from the original on 2013-01-03.
- [8] Halls. "Ideal Weight and definition of Overweight". *Moose and Doc*. Archived from the original on 2011-01-26.
- [9] Frank, DL., Schmid, LT., Sallis, FJ., Chapman, J., and Saelens,EB., "Linking Objectively Measured physical Activity with Objectively Measured Urban Form Findings from SMARTRAQ". *American Journal of Preventive Medicine*, Vol. 28: 117-125,2005.
- [10] Rolls, BJ. "The Supersizing of America: Portion Size and the Obesity Epidemic," *Nutrition Today*, 38, No. 2: 42–53, 2003.
- [11] Carnegie, MA., Bauman, A., Marshall, AL., Mohsin, M., Westly-Wise, V., and Booth, ML." Perception of Physical Environment, Stage of Change for Physical Activity, and Walking among Australian Adults", *Research Quarterly for Exercise and Sport*, 146-155,2002.
- [12] Centers for Disease Control and Prevention (2011) Physical activity and fitness. *Healthy people 2020: Physical activity*.
- [13] Saelens, BE., Sallis, JF., Black, JB., and Chen, D., "Neighbourhood-based Differences in Physical Activity:An Environment Scale Evaluation", *American Journal of Public Health*, Vol.93: 1552-1558, 2003
- [14] Norman GJ, Nutter SK, Ryan S, Sallis JF, Calfas KJ, and Patrick K. "Community Design and Access to Recreational Facilities as Correlates of Adolescent Physical Activity and Body-Mass Index", *Journal of Physical Activity and Health*, Vol. 3: 118-128, 2006.
- [15] Smith, Scott (8 April 2013). "Determining Sample Size: How to Ensure You Get the Correct Sample Size | Qualtrics". *Qualtrics*. Retrieved 15 November 2016.

- [16] Stigler, Stephen M. (1989). "Francis Galton's Account of the Invention of Correlation". *Statistical Science*. 4 (2): 73–79. *JSTOR* 2245329. doi:10.1214/ss/1177012580.
- [17] "Physical status: The use and interpretation of anthropometry" (PDF). WHO Technical Report Series. Geneva, Switzerland: World Health Organization. 854 (854): 1–452. 1995. PMID 8594834. Archived (PDF) from the original on 2007-02-10.

## BIOGRAPHIES OF AUTHORS



Md. Nofayer Haque M.S. Student, Department of Statistics, Jahangirnagar University, Savar Dhaka-1342. He has finished his M.Sc and B.Sc degree from Jahangirnagar University in 2017 and 2016 with major in Statistics. His research interest focus on Artificial neural networks research based problem, Artificial intelligence, Machine Learning, statistical learning methods, public health and biostatistics research based problems.



Muhammad Mahady Hasan M.S. Student, Department of Statistics, Jahangirnagar University, Savar Dhaka-1342. He has finished his M.Sc and B.Sc degree from Jahangirnagar University in 2017 and 2016 with major in Statistics. His research interest focus on Response surface modeling, Design of experiment, statistical learning methods, public health and biostatistics research based problems.



Azizur Rahman, Assistant Professor, Department of Statistics, Jahangirnagar University, Savar Dhaka-1342. He has finished his MPhil and M.Sc degree from Jahangirnagar University in 2017 and 2010 with major in Statistics. His research interest focus on statistical theory and inference, statistical learning methods, public health and biostatistics research based problems.